# Astronomical Spectroscopy: Detecting light with a CCD

October 14, 2014

Jung Lin (Doris) Lee
dorislee@berkeley.edu
Group partners: Jennifer Ito, Manuel Silvia
Prof. James Graham, UGSI Heechan Yuk, Isaac Domagalski

## Abstract

Spectroscopy is an essential tool in observational astronomy for examining properties of distant sources. In this experiment, we observe the spectral line signatures of different everyday light sources. We examine the hardware of the Ocean Optics USB 2000 spectrometer and consider how physical phenomenon such as thermal noise could affect the quality of our data. By taking dark frames, we use the resulting mean and variances to quantify the read noise and Poisson error of the instrument. Then, two different data-smoothing techniques are employed to eliminate tiny fluctuations in the intensity values so that they are not identified as important spectral line signature. Even though data smoothing is supposed to improve the accuracy of the centroid-finding algorithm, we found that our automated centroid-finding algorithm misses a few important spectral line features. Therefore, using the centroid values obtained from the estimation method instead, we conduct a linear least squares fit with the known neon and mercury wavelength values. By comparing the residual plots and the sum of residuals from the linear and quadratic fits, we conclude that a linear fit is good enough for our experiment's data and a higher-order description for wavelength calibration is not necessary.

## 1. Introduction

Optical spectroscopy provides important information about the chemical composition observational targets since astronomers often do not have access to physical samples of their targets for conducting chemical analysis. Detailed insight into the properties of planetary bodies, radiating black-bodies, and their surrounding medium can be derived from atomic absorption and emission features on a spectrum.

Early models of spectrometers consist of prisms that dispersed light into different components . However, most modern day spectroscope uses finely-grated slits instead of prisms to minimize absorption so that faint sources, common to astronomical observation, can be better observed. Moreover, the use of optical fiber plates for simultaneous acquisition of multiple spectra enables cosmological studies such as Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al.) to examine higher redshift quasars and luminous red galaxies.

In this report, we present the internal workings of a spectrometer and how that effects the raw, unprocessed spectra obtained. In section 2, I will present the Ocean Optics USB 2000 spectrometer hardware used for this experiment and a qualitative overview of the data resulting from different sources. In section 3, I will detail the various procedures that were used to eliminate noise and variations associated with the spectrometer's hardware setup. In addition, I will show the Poisson nature of these systematic effects. Finally, section 4 explains how we converted the pixel values recorded by the spectrometer to the more hardware-independent measurement of wavelength by comparing our data with given atomic line features referenced from NIST[1].

---

[1]National Institute of Standards and Technology Atomic

## 2. Hardware and Experiment

### 2.1. Ocean Optics USB 2000 spectrometer

A spectroscope is an instrument used for analyzing components of an electromagnetic wave over some specific wavelength. In a typical spectroscopy setup, the incoming light is collimated by a lens and goes through a diffraction grating that disperses the beam into component of different wavelengths. Then, the dispersed light is detected by the CCD and voltage readout corresponding to the intensity value incident on each pixel is converted into digital units and stored for further analysis.

The spectrometer that we are using for this experiment is an Ocean Optics USB 2000 spectrometer, with a 2048-pixels linear CCD. Each pixel in the CCD spans $14\mu m \times 200\mu m$.

### 2.2. Qualitative findings for different sources

**Color.** In astronomical spectroscopy, color (i.e. the range of wavelength that the peak falls near) yields information about the temperature of a stellar population, which can then serves as an indicator for its age. To explore the color wavelength response of the spectrometer, we place red, green, orange, and blue colored paper in front of a very bright LED flashlight. Even though there was transmission loss through the filter, the absorption should be fairly equal for all wavelengths. Therefore we can still clearly see from Fig. 1 the wavelength correspondence of each color in the visible wavelength.

**Voltage.** The desk lamp is attached to an external potentiometer which allows us to vary the voltage of the source. Since to the relationship between voltage and current is described by Ohm's law, the decrease in voltage is directly proportional to the decrease in current, which leads to a dimming of the lamp. Another way is to consider the power dissipation by the bulb as described by Eq.1:

$$P = IV = \frac{V^2}{R} \tag{1}$$

where P is the power dissipated by the lamp, I is current through the lamp, V is the voltage across the circuit (externally modulated by the potentiometer) and
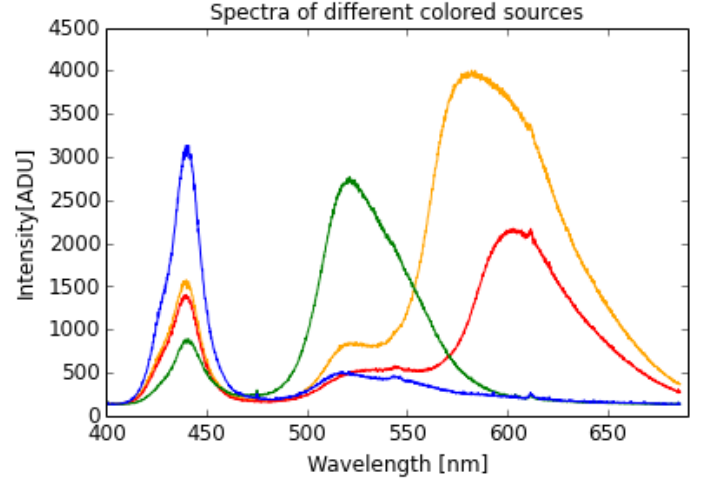
Figure 1: By adding a color filter in front of a radiating body, we found that the lines peaks around the color's corresponding wavelength in the visible spectrum. The difference in relative intensity is irrelevant, as it may simply be due to how close we are holding the source to the entrance slit of the spectrometer and how much light the colored-marker ink is blocking out. The pixel-to-wavelength conversion was done according to the manufacturer's wavelength calibration coefficients.

R is the resistance of the circuit which in this case is fixed. Therefore, decreasing the voltage results in a decreases in power dissipation.

If we approximate the desklamp source as a radiating blackbody, Wien's law states that its wavelength distribution for different temperature should be approximately the same shape but with a displaced wavelength as shown in Eq.2:

$$\lambda_{max} = \frac{b}{T} = \frac{2.897 \times 10^3 m \cdot K}{T} \tag{2}$$

where $\lambda_{max}$ is the peak wavelength for the distribution, T is the temperature of the radiating blackbody, and b is the Wien's displacement constant. Since Eq. 1 shows a lower power dissspation for smaller input voltage, it means that the lamp's bulb is cooler in temperature than if it was at a higher voltage setting. This is why the peak wavelength of the lower-voltage source distribution is longer than the higher-voltage source.

**Intensity.** The sunlight spectra was taken by an optical fiber that lead from spectrometer's entracnce slit to an adjustable stand to minimize the variation of the , however external factors such as cloud and shadows still varied the source brightness. Also, despite the
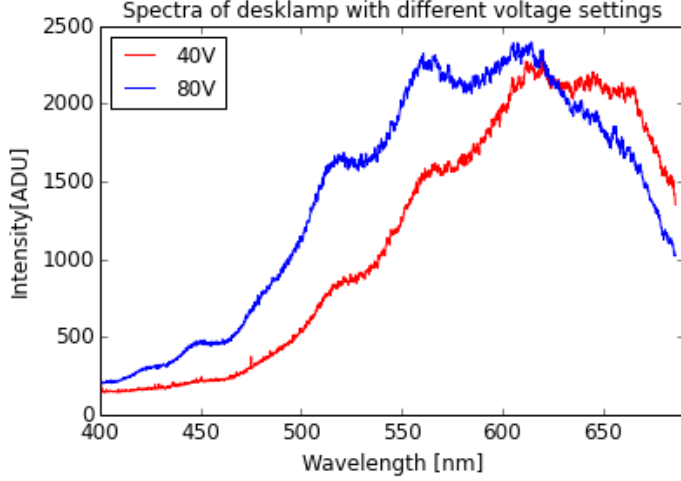
Figure 2: Changing the input voltage of the source affects the peak wavelength of the distribution.

use of minimum integration time (3ms) for data acquisition, the pixels still saturate by the high intensity of sunlight as shown in the cutoff in Fig.3. Averaging the data does not alleviate the effect since this feature persists throughout the whole duration of taking the 1000 samples.
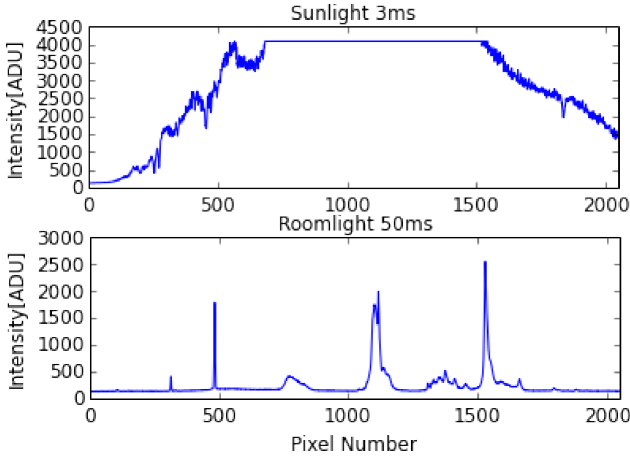


Figure 3: Top: Saturation of the CCD pixels results in the plateau of the spectra. Bottom: Mercury lines in roomlights and spectra obtained from other sources did not exhibit this plateauing effect since none were of such high intensity.

## 3. Noise and Systematic Effects

### 3.1. Natural Broadening Effect

As seen in the mercury lines in Fig.3, the spectral lines do not resemble perfectly sharp Dirac Delta functions. This broadening effect is due to several physical phenomena. One reason for the natural line width is the measurement uncertainty inherent from quantum mechanics. Another more dominant effect is due to the Doppler effect of atoms moving with thermal velocity while they are emitting the electromagnetic wave recorded in the spectra. The distribution of the atomic thermal velocity is governed by the Boltzmann distribution. This Doppler-shifted velocity distribution propagates to our intensity measurement, which can then be rearranged into a Gaussian form. The broadening effect is characterized by the variance of this new distribution.

### 3.2. Dark Counts

We took set of 3ms integration time bias frames by placing the red cap on the spectrometer and in a black bag. This integration time was chosen as close to zero as possible to minimize the tails of the Gaussian on the two ends. The 3ms integration time is the shortest integration time accessible for data acquisition using the SpectraSuite software. These dark count measurements can be used to subtract out the per-pixel noise in the experiment's data.
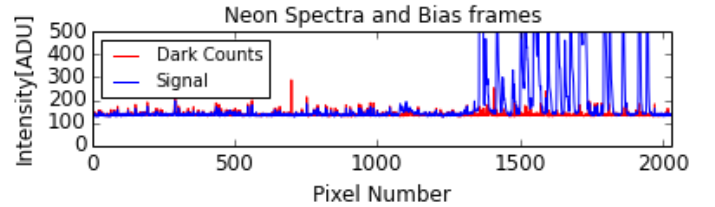


Figure 4: A spectrum of neon zoomed in at lower intensity value to show the relative intensity of the signal versus the noise from the dark counts.

In the Ocean Optics USB 2000 spectrometer, the first 24 pixels are covered up during each taking of the spectra. These dark pixels could also serve as a way to correct for the noise level offset. In real astronomical observations, the method of dark subtraction as shown in Fig.5 and the use of overscan strips by adding additional pseudopixels are two common ways of bias calibration on imaging data. Although our spectrometer has a linear CCD array as a detector, CCD for imaging telescope often has a plane array of pixels. In those cases, averaging 10 or more bias frames is often a better bias-correction technique since it reveals any underlying 2D bias patterns present in the image.
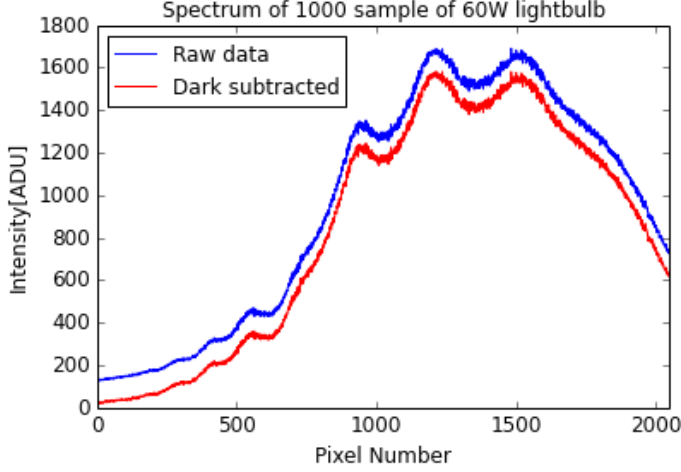
Figure 5: A spectrum of 60W lightbulb before and after pixel-by-pixel dark subtraction. The small fluctuations present in both spectra may be due to flat field variations across the CCD's linear array of pixels.

### 3.3. Poisson nature of obtained data

We chose to conduct the statistical analysis of noise property on the 60-Watt lightbulb data since the variance is less likely to be dominated by fluctuations in the source's illumination brightness. A Gaussian is fitted onto the histogram of dark frames in Fig.6 to obtain the value of mean ($\mu$) and variance($\sigma$) of the distribution. The $\mu$ corresponds to the $ADU_0$ in Eq.3 , whereas the $\sigma$ corresponds to $\sigma_{ADU_0}$, the error on the dark counts. This variation on the dark counts comes from the non-uniform quality of the pixel as seen in Fig.6. But since $\sigma_{ADU_0}$ is relatively small compared to $ADU_0$, it is negligible when considering its per-pixel additive effect on the raw data.

$$ADU = g\frac{Ne}{C} + ADU_0 \qquad (3)$$

Analog digital units (ADU) is the value of the digital measurement that is converted from the raw electron count in the circuit; C is the capacitance where the total electric charge from N number of electrons with charge e is deposited and g is simply a proportionality constant for the conversion.

Since the CCD detector is essentially doing photon counting during the course of an experiment, the CCD noise is largely governed by Poisson statistics. By conducting the method of error propagation on Eq.3 and assuming the Poisson and read noise is un-
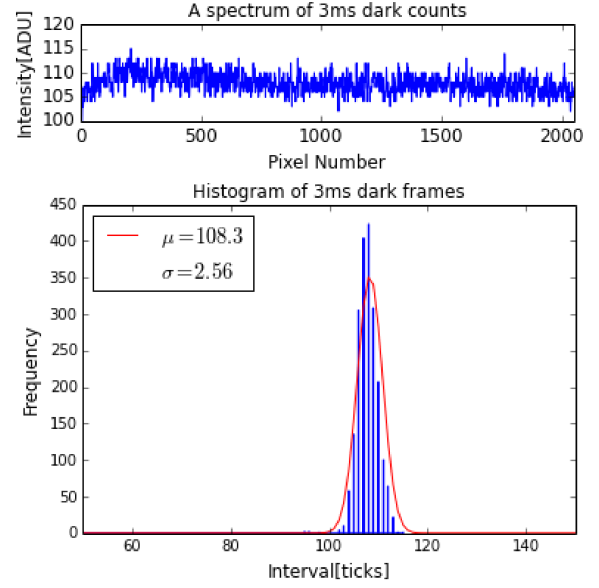


Figure 6: Top panel shows a spectrum of dark frames that exhibits the non-uniformity of each pixel in the linear CCD array. We can see that even when no light is entering the spectroscope, there are non-zero intensity values due to the effects of dark current and bias offset. The spectrum roughly corresponds to the bias value ($ADU_0$) obtained from the $\mu$ of 108 ADU shown in the bottom panel.

correlated, we obtain :

$$\sigma_{ADU}^2 = \left(\frac{\partial ADU}{\partial N}\right)^2 \sigma_N^2 + \left(\frac{\partial ADU}{\partial ADU_0}\right)^2 \sigma_0^2 \qquad (4)$$

where $\sigma_0$ is the read noise associated with each measurement and $\sigma_{ADU}$ is the Poisson noise which is equal to N (due to notable $\sqrt{N}$ Poisson standard deviation) . Then, writing this in linear form by substituting Eq.3, we get :

$$\sigma_{ADU}^2 = \frac{ge}{C}(ADU - ADU_0) + \sigma_0^2 \qquad (5)$$

The list of mean and variances are computed by stacking together in full datasets of spectra and retrieving the intensity value for each pixel over all the dataset. Then, using the 1000-sample slice of intensity values for the same pixel to compute the mean and standard deviation. We observe the linear relation between $\mu$ and $\sigma$ due to the Poisson noise.

### 3.4. Other possible source of error

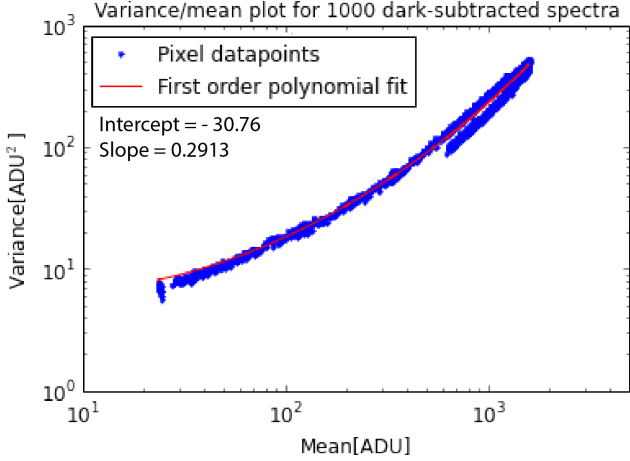Since our spectrometer operated at room temperature, a source of error in our experiment is dark

4

Figure 7: The linear relation of variance and mean can be observed in this plot. Since we are plotting the mean of dark-subtracted signal (ADU-$ADU_0$). The slope from the linear fit is equal to the the gain and the intercept is the (read noise)[2].

current. This effect is due to thermal noise in the spectrometer's readout electronics and in the CCD. Most modern day telescopes have a dedicated dewar that cools their CCDs to minimize the production of thermal electrons. This is especially important when the observing target is faint, (i.e. low signal-to-noise ratio), as the detector has no way to distinguish between thermal electrons from real astronomical photoelectrons.

Another source of error in our experiment is multiple mixed sources present when taking data. We took preventive measures when taking our data for neon and coarsely covered up the gap between the slit entrance and the neon bulb with its transparent gap removed. This problem could have been completely resolved if data was taken in the dark so that no slight traces of atomic feature of radiating source in the background can be seen in the pure neon sample used for wavelength calibration.

## 4. Data Reduction

Now we have examined the systematic effects that affects our data, we process the raw data obtained from the spectrometer in order to find an equation that converts the pixel values to wavelengths.

### 4.1. Data Smoothing

Smoothing is a data reduction technique that removes noise and minor pixel-to-pixel variations in the data. We employ both the boxcar and averaging technique in our datasets and compare their results.

***Boxcar Smoothing***. Our boxcar smoothing algorithm takes the values of adjacent pixel and averages them over some user-defined interval. Since the quality of each pixel is inconsistent, boxcar smoothing compensate for the pixel-by-pixel variations. However, by lowering the signal-to-noise ratio to reduce small-scale features, we also lower the resolution of the delta-function-like peak of important spectral features. When the interval is chose incorrectly, datasets processed in such manner can have adverse effects on centroid value calculation as shown in Fig.8 .
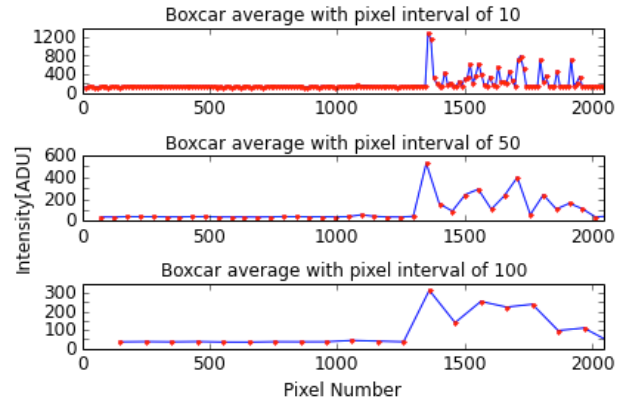


Figure 8: The red marker denotes the interval that we are averaging over. As the number of pixels in an averaging interval increases, the shape of the line and the location of its extrema is obscured by the lack of datapoints defining it.

***Average multiple spectra samples***. Since the boxcar method produced undesirable results, we resorted to the conventional averaging of data by summing over the intensity values of each pixel of the same experiment then scaling the result by the number of datasets. This is a step towards finding a functional form to fit the data since it makes the data less discrete and more continuous like a function.

### 4.2. Centroid-Finding Methods

The centroid for a spectrum is a weighted average of position and intensity computed by Eq.6. The computation itself is trivial but the hard part is finding the interval that we are taking the summation over. We employ two different techniques in finding

5

the centroid by determining the appropriate interval to compute the sum.

$$\text{Centroid} = < x_i > = \frac{\sum\limits_{i=1}^{N} x_i I_i}{\sum\limits_{i=1}^{N} I_i} \qquad (6)$$

where x is the pixel position value and I is the intensity.

***Automated centroid algorithm.*** We attempted to create an automated centroid algorithm by imposing two conditions to select for the centroid of interest that correspond to atomic signature spectral lines. The selected data range used for centroid calculation must begin at a local minimum and contain one local maximum and then end with another local minimum. There can not be any extra extrema that lies within that range other than the 3 specified. Since a large number of datapoints satisfies this condition, we needed a quantitative measure of how "important" each spike is.

For this, we use the intensity difference between the consecutive local maximum and minimum and select only the data ranges that have differences greater than 10% of the difference between the absolute maximum and minimum datapoint of the whole dataset. The ad hoc 10% criterium was chosen to so that the bias offset was considered in . The 10% was chosen because of the quality of the centroid result that it produces. The local extrema was found by comparing consecutive data points and testing a change in the boolean condition of increasing or decreasing.

The algorithm was not very robust as it is unable pick out spectral lines that lied between consecutive extrema, such as the ones near pixel 1115 and 1374 in Fig.10. Despite the attempts to vary the percentage cut, the program cannot identify features that showed relative differences, so it tends to overlook single, short, sharp spikes. The variability in the importance of a signature line is hard to account for by measuring only the absolute differences.

***Estimation method.*** If we approximate the shape of a single atomic line feature as a Gaussian, an isolated data range of plus-or-minus one sigma around the mean accounts for about 68% of the area under
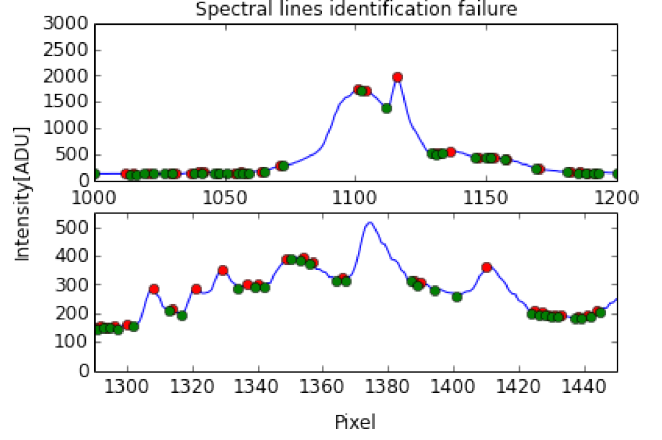


Figure 9: Top: The closely spaced extremum at around the 1100th pixel results in a small difference value between the local minimum and maximum. Therefore, it was not detected by the algorithm. Bottom: The extremum finding algorithm did not detect the maximum at around pixel 1380 in its first step.
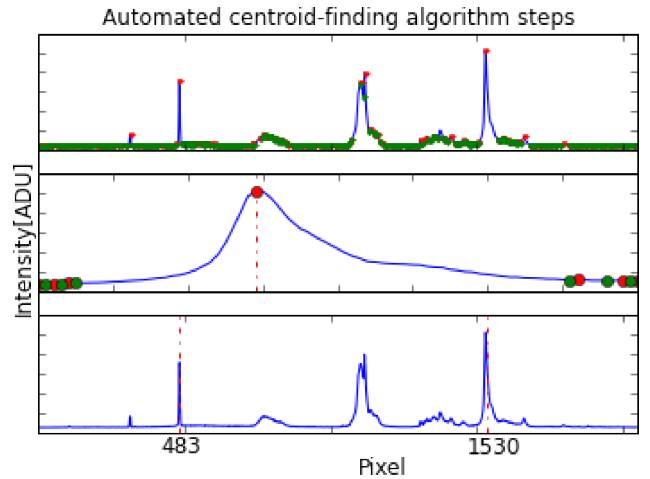


Figure 10: Graphical representation of how the automated algorithm works. Top: Identifying all the extremum of the dataset. Red marker denotes local maximum; green denotes local minimum. Middle: Applying criteria to select suitable ranges as shown in middle plot, using this sliced dataset to obtain the centroid values (dashed). Bottom: Resulting centroid list misses some important spectral feature picked out by the estimation method.
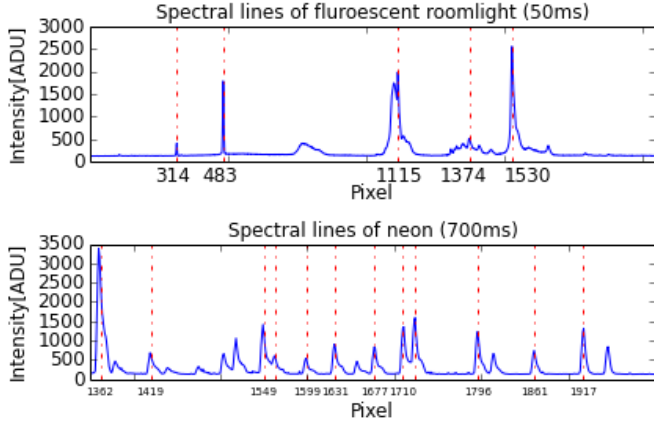
6

Figure 11: Final spectral lines for neon and mercury obtained by estimation method, used for wavelength calibration.
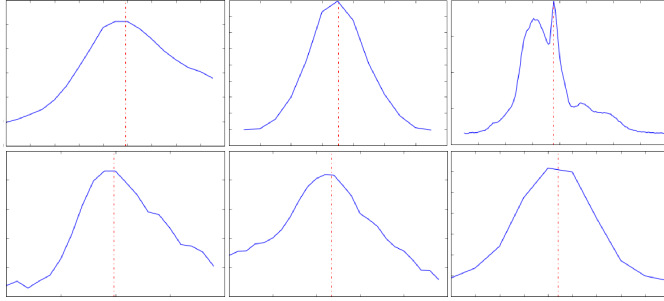


Figure 12: Windows selected by visual inspection before quantifying the range with one sigma.

a Gaussian based on the 68-95-99.7 rule (Wall and Jenkins, 2003). We manually select data range based on visually inspecting where a spectral line lies as shown in Fig.12. Then, we calculate the centroid on $\pm 1\sigma$ of the maximum peak in the selected data range. The 1-$\sigma$ window provides a quantitative way of defining the boundaries of the data range that we are taking. Trying to automate this procedure would be challenging because not all the atomic line are symmetric Gaussians so the centroid may be skewed to one side of the peak.

*4.3. Wavelength calibration*

The polynomial nature of our fitting model comes from the perpendicular incident ray case of the grating equation. By approximating that the light's wavelength is small relative to the groove spacing of the diffraction slit, a Taylor expansion can boiled down the trigonometric equation to a polynomial with higher order terms. In this case, the fitting parameter is the polynomial coefficients The residual is defined
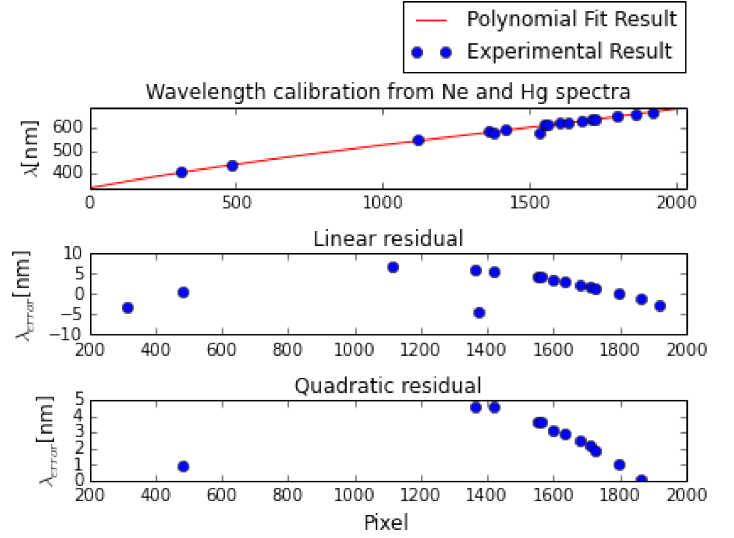


Figure 13: Top panel shows the result of a linear least square fit on the centroid values obtained from the neon and mercury spectra. The middle and bottom panel is residual plot from comparing against the linear and quadratic fits respectively.

|  | Linear | Quadratic | Manufacturer |
|---|---|---|---|
| $a_0$ | 356.326 | 351.556 | 344.311846 |
| $a_1$ | 0.16380 | 0.175133 | 0.19743726 |
| $a_2$ | — | $-5.10 \times 10^{-6}$ | $-1.4913 \times 10^{-5}$ |
| $\sum r_i$ | $-2.16 \times 10^{-12}$ | $-9.265 \times 10^{-12}$ | -2.292 |

Table 1: The wavelength calibration coefficients obtained from experimental data. $r_i$ is the residual calculated for each datapoint. Note that two outliers were removed when computing manufacturer's total residuals, in order to show how close does the manufacturer's calibration fit with the majority of the datapoints we obtained.

as $y_i - f(x_i)$ where $y_i$ is the centroid value in pixels from experimental result and f is the polynomial fitting model. We plot the residual in order to examine whether there are any patterns present in the residual which would suggest a need for a higher order fit.

## 5. Conclusion

In this experiment, we measured the effect of different light-emitting sources and draw a qualitative relationship between spectral responses to the underlying physics in play. By decreasing the voltage parameter of an incandescent source, we found that the peak wavelength is red shifted due to the lower temperature from less power-dissipation. The arrangement of colored filters exhibits higher intensity peaks
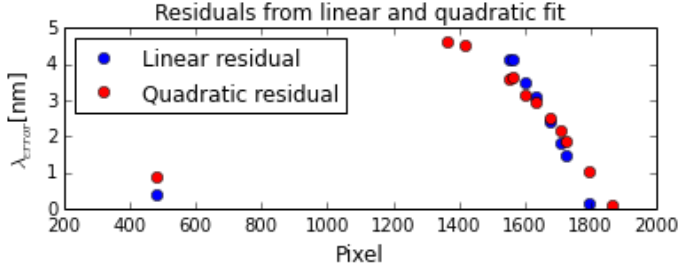
Figure 14: Above is an overlaid plot of the linear and quadratic residuals. It shows that there is not much improvement in using a higher order polynomial model than the linear one, which is also evident from the sum of residuals computed in Table1. This makes sense because the polynomial coefficient that we obtained for the quadratic term is $-5.101 \times 10^{-6}$, which is an order of magnitude smaller than the quadratic coefficient in the manufacturer's wavelength calibration. This is the reason why the y-axis range of the linear residual plot in Fig.13 is almost the same as the range in the quadratic residual plot.

to the filter's corresponding wavelength. By examining the hardware setup and the internal workings of a spectrometer, we consider the possible systematic effects on the data taken from the experiment.

During data reduction, the data was smoothed using the boxcar and averaging methods. We find that the averaging method serves this purpose better since it is not subjected to loss of per-pixel detail as much as the boxcar technique does. With the smoothed data, the automated centroid-finding algorithm can better identify centroids of interest in isolated window ranges, as some false-positives (short spiky peaks that do not resemble important spectral features) are eliminated by the smoothing procedure. However, since our algorithm only compares absolute differences in the local extrema of the spectra, it misses some important spectral line features in the processed data. Therefore, we resorted to visually-examining the spectral peaks and defining a window of peak value±1$\sigma$ so that a more accurate centroid can be calculated in the data range.

With the list of centroid values, we conduct linear least square fit on the wavelength value with the spectral line pixel and wavelength values referenced from NIST. Then, we conduct a higher-order quadratic fit. By comparing the sum of residuals from the two fits and examining the residual plots, we find that there was not much of an improvement for going to a higher order fit. Finally, the fitting parameters are the

coefficients used for wavelength calibration, which enable us to express the pixel numbers obtained in our spectra data to the less hardware-dependent values of wavelength.

As we only applied dark frame corrections to our experiment data, one possible extension to this project may be to try conducting a basic flat-field correction to the detector. One way to do this to shine bright light uniformly across the detector to see the response of each pixel. The exposure time needs to be short so that the CCD is not saturated. In doing so, we can also try to measure maximum ranges at which CCD is sensitive and linear.

## 6. References

- Howell, Steve, *Handbook of CCD Astronomy*, 2nd Edition. Cambridge University Press, 2006.

- Haken, Hermann and Wolf, Hans Christolph, *The Physics of Atoms and Quanta*, Springer-Verlag, 1996.

- Wall, J. V. and Jenkins, C.R., *Practical Statistics for Astronomers*, Cambridge University Press, 2002.

- Hecht, Eugene, *Optics*, Addison-Wesley, 2001.